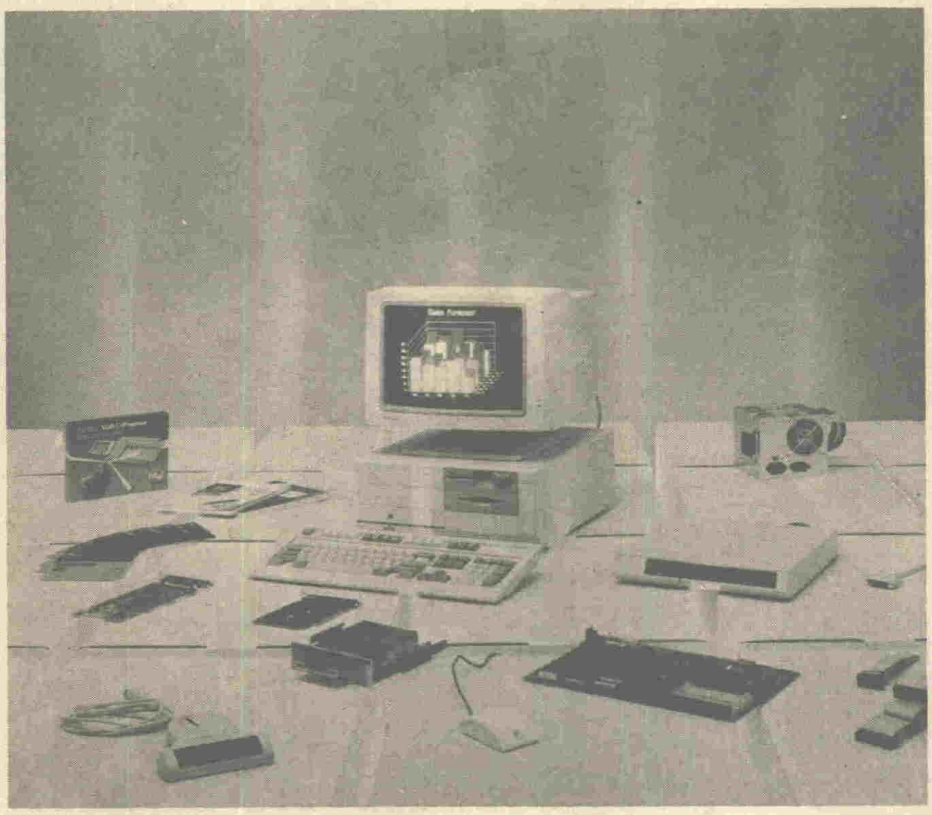
เรื่องน่าอ่าน

เอนก ศิรจริยาภิวัฒน์

# Cache
## คำ ๆ นี้มีอะไรซ่อนอยู่หรือ



***คุณแน่ใจแล้วหรือว่า***
***คุณเข้าใจ***
***หน่วยความจำแคช***
***ได้ลึกซึ้ง***
***และดีพอสมควรแล้ว***

ทุก ๆ คนมักจะพูดถึงคำว่า เมกะเฮิรตซ์ (MHz) และ Million instruction per second (ทำงานได้กี่ล้านคำสั่งต่อวินาที) การพูดถึงแต่ความเร็วของซีพียูนั้นไม่ได้เป็นจุดสำคัญในการวัดประสิทธิภาพหรือความสามารถของระบบ แม้ว่าเราจะมีซีพียูที่วิ่งได้เร็วดังจรวดอย่าง 80486 ขนาด 50 เมกะเฮิรตซ์ แต่ถ้ามีระบบการส่งถ่ายข้อมูลไม่ดีพอ ซีพียูดังกล่าวก็เปรียบเสมือนคนพิการที่เดินไปไหนมาไหนไม่สะดวกนั่นเอง ซีพียูตัวที่เร็วที่สุดในโลกก็สามารถวิ่งได้เร็วสูงสุด แค่ความเร็วสูงสุดในการส่งถ่ายข้อมูลเท่านั้น ปัจจุบันปัญหาที่พบกับซีพียูที่มีความเร็วสูง ๆ มีอยู่ 2 ประการก็คือ ประการแรกหน่วยความจำที่มีอยู่ปัจจุบัน เริ่มมีความเร็วไม่เพียงพอสำหรับซีพียูรุ่นใหม่ (จำเป็นต้องใส่ wait state เพื่อให้หน่วยความจำทำงานทันซีพียู) ปัญหาประการที่สองก็คือ หน่วยเก็บข้อมูลถาวร (secondary storage) เช่น ฮาร์ดดิสก์ หรือฟลอปปี้ดิสก์ มีการทำงานที่ช้ากว่าซีพียูมาก

## ทางแก้ปัญหาทางหนึ่ง

แคช (Cache) เป็นทางแก้ปัญหา 2 ประการดังกล่าว แต่บางครั้งก็ทำให้เราสับสนบ้างเมื่อพูดถึงคำว่าแคชเฉย ๆ เหมือนกับเราพูดถึงแอปเปิ้ลและส้มในเวลาเดียวกัน ซึ่งจริง ๆ แล้วขึ้นอยู่กับว่าเรากำลังพูดหรือนำเสนออยู่ในเรื่องใด (หน่วยความจำหรือดิสก์)

– Memory caching เป็นส่วนช่วยให้หน่วยความจำที่ทำงานช้า (DRAM) ไม่ต้องถูกใช้งาน (access) บ่อยครั้งเกินไป โดยการใช้หน่วยความจำที่ทำงานได้เร็วกว่า (SRAM) มาขั้นกลางระหว่างซีพียูและ DRAM ค่อยเก็บข้อมูลที่ถูกใช้บ่อย ๆ ไว้ในตัวแคชเอง เมื่อซีพียูต้องการข้อมูลที่ตรงกับข้อมูลที่มีอยู่แคชก็จะได้ข้อมูลรวดเร็วขึ้น สำหรับ Caching memory นี้เราจะพบอยู่บนเมนบอร์ดของระบบ (หรือภายในซีพียูบางตัว หรือบางรุ่น เช่น 80486) ซึ่งแคชประเภทนี้จะมีผลกับการเข้าถึงข้อมูล (access) ของดิสก์น้อย (ไม่มีผลอย่างเห็นได้เมื่อมีการใช้ข้อมูลบนดิสก์)

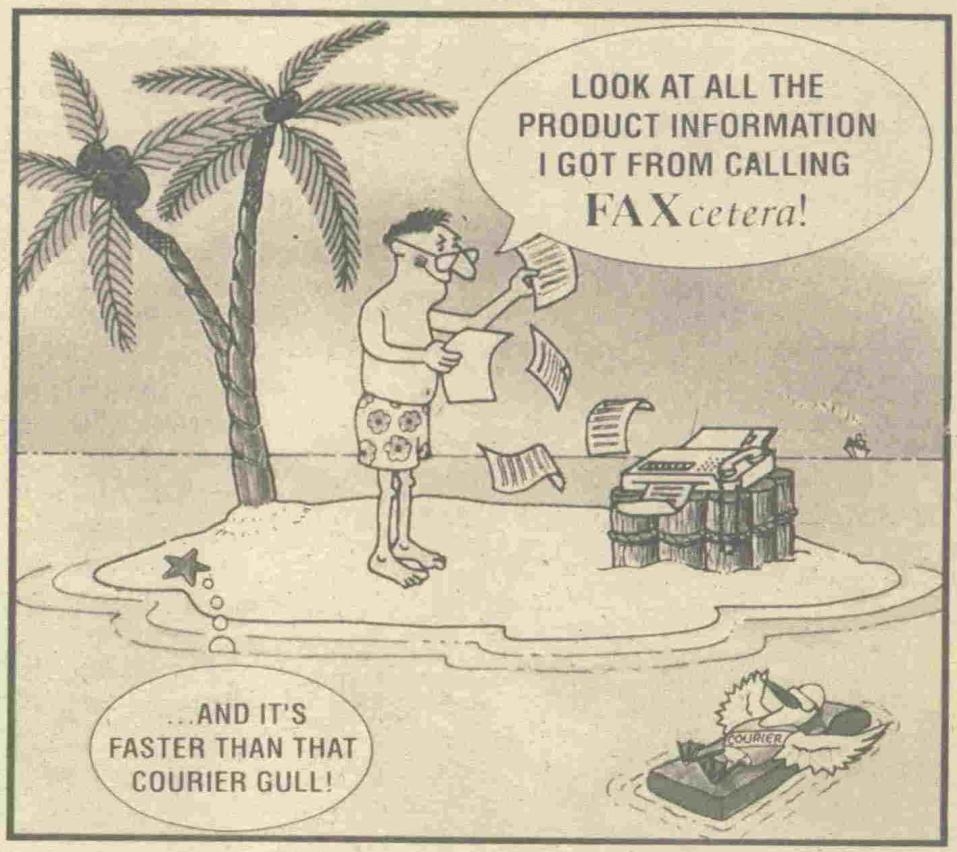
- Disk caching ทำได้ 2 ลักษณะ คือ ซอฟต์แวร์และฮาร์ดแวร์ ตัวอย่างซอฟต์แวร์ก็ได้แก่ โปรแกรม PC-Kwik ของมัลติซอฟต์หรือ Ncache ของนอร์ตัน ซอฟต์แวร์ดิสก์แคชนี้จะมีราคาถูกกว่าฮาร์ดแวร์แคช และสามารถทำงานได้

ค่อนข้างเร็วกว่าฮาร์ดแวร์แคช ทั้งนี้เพราะซอฟต์แวร์แคชใช้หน่วยความจำหลัก ซึ่งมีความใกล้ชิดกับซีพียูมากกว่าหน่วยความจำที่อยู่บนดิสก์คอนโทรลเลอร์ แต่ซอฟต์แวร์แคชจะมีโอกาสทำให้ระบบแฮงได้ง่ายกว่าฮาร์ดแวร์แคชมาก ทั้งนี้เพราะซอฟต์แวร์แคชอาจถูกทำลายไปเมื่อไรก็ได้ นอกจากนั้นฮาร์ดแวร์แคชจะทำให้ซีพียูไม่ต้องแบ่งเวลามาทำงานเกี่ยวกับแคช (ซีพียูจะไม่รู้ว่ามีแคชอยู่หรือไม่? ซึ่งต่างกับซอฟต์แวร์แคชที่ต้องใช้เวลาของซีพียูมาทำหน้าที่บริหารแคช)

## บางสิ่งบางอย่างที่ดูเหมือนจะเป็นแคช

ไม่ใช่ว่าหน่วยความจำที่นำมาใช้เพื่อช่วยการทำงานของดิสก์จะเป็นแคชไปทั้งหมด อย่างเช่น บัฟเฟอร์ของดอส ซึ่งใช้วิธีการแบบ first in/first out ซึ่งเป็นวิธีการที่ง่ายเกินไปที่จะเรียกว่าแคชได้ เราจะบอกว่าแคชต่างกับบัฟเฟอร์ได้ด้วยกฎเกณฑ์ข้อใดนั้น บางทีก็หากฎเกณฑ์ตายตัวไม่ได้ แต่ถ้าจะตอบแบบกำปั้นทุบดิน เราก็บอกได้ว่าแคชมีไว้บริหารข้อมูล (manage data) แต่บัฟเฟอร์มีไว้เก็บข้อมูล (store data) เท่านั้น และนี้เป็นคำตอบอย่างหนึ่งที่ว่าการ์ดควบคุมดิสก์ (disk controller) ที่มี SRAM อยู่บนการ์ด ไม่จำเป็นต้องเป็นคอนโทรลเลอร์ที่มีแคชอยู่ในตัว ทั้งนี้เพราะ SRAM ที่มีอยู่บนการ์ดอาจทำหน้าที่เป็นบัฟเฟอร์ความเร็วสูงเท่านั้น คอนโทรลเลอร์บางตัวก็เพิ่มประสิทธิภาพการทำงานขึ้นไปอีก โดยการอ่านข้อมูลล่วงหน้า (read ahead capabilities) มาเก็บไว้บนคอนโทรลเลอร์ ซึ่งดิสก์คอนโทรลเลอร์แบบ IDE ก็ใช้เทคนิคนี้

เทคนิคการอ่านข้อมูลล่วงหน้านี้ จะใช้วิธีการอ่านเซกเตอร์ทั้งหมดของแทร็กมาเก็บไว้บนคอนโทรลเลอร์ (แม้ว่าโปรแกรมอาจสั่งให้ติดต่อกับดิสก์เพียง 1 หรือ 2 เซกเตอร์) โดยการทำนายว่าการติดต่อกับดิสก์ปกติ มีความน่าจะเป็นไปได้สูงว่าจะติดต่อเซกเตอร์ที่ต่อเนื่องกันไป ซึ่งถ้าเป็นเช่นนั้น จะทำให้การติดต่อดิสก์เป็นไปได้เร็วยิ่งขึ้น (ตัวอย่างเช่น WD1009V-SE2 ของ Western Digital และ ACB-2322D ของอะแดปเท็กซ์)




## การออกแบบและขนาดของแคช

การออกแบบแคชที่สามารถทำงานได้ฉลาดและทันท่วงทีนั้น ผู้ออกแบบจะต้องเผชิญปัญหาต้องพิจารณาการทำงานหลาย ๆ อย่างของแคช เราจึงพบว่ามีวิธีการแตกต่างกันอยู่มากมายหลายวิธีในการออกแบบโปรแกรมแคชและฮาร์ดแวร์ (ติดตามอ่านการทำงานและวิธีการของแคชได้ในฉบับหน้า) ซึ่งจะพบว่าตอบคำถามได้ไม่ง่ายเลยว่าจะสร้างแคชอย่างไร ตัวอย่างเช่น จะออกแบบให้มีขนาดใหญ่ไว้ก่อน ซึ่งยังเป็นวิธีที่ไม่ได้ถูกต้องเสมอไป ทั้งนี้เมื่อเราเริ่มใช้แคชในขั้นแรก เราจะพบว่าเราสามารถเข้าถึงข้อมูลได้เร็วขึ้น แต่เมื่อแคชมีขนาดใหญ่ขึ้นถึงจุดหนึ่ง เราจะพบว่าการเข้าถึงข้อมูลนั้นจะไม่เร็วขึ้น และเมื่อแคชมีขนาดใหญ่เกินค่าดังกล่าว จะพบว่าการเข้าถึงอาจมีค่าช้าลง ทั้งนี้เพราะว่าข้อมูลทั้งหมดไม่ได้อยู่ในแคช เมื่อเราต้องการข้อมูล (ซึ่งข้อมูลนั้นไม่ได้อยู่ในแคช) จะทำให้เสียเวลาในการค้นหาข้อมูลในแคชไปอย่างเปล่าประโยชน์

# ต้นฉบับไม่มีหน้า 243

# No This Page 243 In Original